# Bayesian Experimental Design for Another Bayesian's Analysis (or a Frequentist's)

Chin-Pei Tsai* and Kathryn Chaloner*

Tech Report #630
March 24, 1999

# Bayesian Experimental Design for Another Bayesian's Analysis (or a Frequentist's)

Chin-Pei Tsai* and Kathryn Chaloner*

April 1, 1999

**Abstract**

Etzioni and Kadane (1993) developed an approach to experimental design for another decision maker's analysis. Their approach is that of a single decision maker, the designer, designing the sample optimally to estimate with the suboptimal estimate of the evaluator's decision. Lindley and Singpurwalla (1991, 1993) developed a similar approach for acceptance sampling and life testing where a seller decides on a sample for a buyer.

Our alternative approach assumes that the designer designs the experiment to optimize the inference of the evaluator. The quantity that the designer minimizes is the designer's prediction of the Bayes risk of the evaluator; in contrast, in Kadane and Etzioni's approach, the designer minimizes the designer's prediction of the Bayes risk of the *designer* using the evaluator's estimator. An important special case is when the evaluator uses a vague or improper prior distribution: this gives an approach to Bayesian design for a frequentist analysis. Examples of design for normal linear models, nonlinear models and normal, Bernoulli and Poisson sampling are given. In the normal linear regression model, the standard classical optimal design theory can be given a Bayesian justification.

*Chin-Pei Tsai is a graduate student and Kathryn Chaloner is a Professor at the School of Statistics, 1994 Buford Ave., University of Minnesota, St. Paul, MN 55108. This research was partly supported by a grant from the National Security Agency.

# 1. INTRODUCTION AND MOTIVATION

Lindley and Singpurwalla (1991, 1993), Etzioni and Kadane (1993) and Lodh (1993) developed an approach to Bayesian design for another Bayesian's analysis. The two Bayesians, the designer and the evaluator, each have their own prior beliefs about a quantity of interest. The designer can design the experiment but the evaluator's estimator must be used. The designer chooses the design and the evaluator chooses an estimator. The designer is the decision maker and views the problem as using a suboptimal estimator. The designer therefore chooses the design to minimize the designer's Bayes risk for that estimator. We call this approach "Approach 1".

Let $\pi_e(\theta)$ and $\pi_d(\theta)$ be the prior distributions used for estimation and design respectively. The corresponding posterior distributions for a sample $\boldsymbol{y}$ of $n$ observations are $\pi_e(\theta|\boldsymbol{y})$ and $\pi_d(\theta|\boldsymbol{y})$. Let $p_d(\boldsymbol{y})$ and $p_e(\boldsymbol{y})$ be the marginal, or predictive distributions, of the designer and evaluator for the data. Approach 1 minimizes $E^{p_d(\boldsymbol{y})}E^{\pi_d(\theta|\boldsymbol{y})}L(\theta,a)$, the designer's predictive risk for the evaluator's estimate, where $a=a(\boldsymbol{y})$ minimizes $E^{\pi_e(\theta|\boldsymbol{y})}L(\theta,a)$.

An alternative viewpoint, motivated in this paper, is that the designer should design to make the evaluator's inferences as precise as possible from the evaluator's viewpoint. The alternative approach, denoted "Approach 2", minimizes the expectation, under the designer's marginal predictive distribution of data, of the evaluator's Bayes risk. That is if $a$ minimizes $E^{\pi_e(\theta|\boldsymbol{y})}L(\theta,a)$, in Approach 2 the design is chosen to minimize $E^{p_d(\boldsymbol{y})}E^{\pi_e(\theta|\boldsymbol{y})}L(\theta,a)$. In Approach 1 the design is chosen to minimize $E^{p_d(\boldsymbol{y})}E^{\pi_d(\theta|\boldsymbol{y})}L(\theta,a)$.

Approach 2 is appropriate when the designer has prior information and is willing to use all the prior information for design but does not want to use all the prior information for analysis. The designer wants the analysis to be

the best analysis possible for a prior distribution which does not represent all the prior information available. This might be the case, for example, if the designer wishes to design an experiment to convince a scientific community which does not share the designer's prior beliefs. The designer, therefore, serves the needs of the evaluator. Another motivation is that the designer is a company with a lot of additional information, but the evaluator is the FDA who will use a maximum likelihood analysis or inferences from a reference prior distribution.

Approach 1 can also be implemented using a non-informative prior distribution but in Approach 1 it is the posterior risk of the *designer* not the posterior risk of the evaluator that is minimized. The differences between Approach 1 and 2 are illustrated with examples.

## 1.1 The Linear Model

Consider a linear model where $\boldsymbol{y}$, given $\boldsymbol{\theta}$ and $\sigma^2$, is a vector of $n$ independent responses with mean $\boldsymbol{X\theta}$ and variance $\sigma^2\boldsymbol{I}$, where $\sigma^2$ is known. Here $\boldsymbol{\theta}$ is a vector of $p$ parameters and $\boldsymbol{X}$ is the design matrix. Suppose the prior distributions are normal conjugate prior distributions. That is the prior distribution are $\boldsymbol{\theta}|\sigma^2 \sim \boldsymbol{N}(\boldsymbol{\mu}_i, \sigma^2\boldsymbol{R}_i^{-1})$, where $\sigma^{-2}\boldsymbol{R}_i$ is a prior precision matrix and $i = e, d$ for evaluator and designer respectively. The posterior mean of the evaluator is $\hat{\boldsymbol{\theta}}_e = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{R}_e)^{-1}(\boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{R}_e\boldsymbol{\mu}_e)$ with posterior variance $\sigma^2(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{R}_e)^{-1}$. Similar expressions represent the posterior distribution of the designer with $\boldsymbol{R}_d$ and $\boldsymbol{\mu}_d$ replacing $\boldsymbol{R}_e$ and $\boldsymbol{\mu}_e$. This is an unusual situation where the posterior variance does not depend on the data $\boldsymbol{y}$. For fixed $n$, therefore, applying Approach 2 for any designer's prior distribution leads to choosing the design to optimize some function of

$\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{R}_e$. For example with square error loss for the estimation of $\boldsymbol{c}^T\boldsymbol{\theta}$, Approach 2 leads to minimizing $\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{R}_e)^{-1}\boldsymbol{c}$, the usual c-optimality for the evaluator's prior distribution. If $\boldsymbol{R}_e \to 0$ the evaluator has an improper non-informative distribution, the evaluator's posterior mean is the least squares estimate and Approach 2 gives classical, non-Bayesian, optimal design (as in Fedorov, 1972, Silvey, 1980, Pukelsheim, 1993, for example) with $D$-optimality or $A$-optimality under suitable loss functions.

Approach 1 views the problem as optimal design for a suboptimal estimate. The designer's posterior will also be normal with posterior mean $\hat{\boldsymbol{\theta}}_d = (\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}_d)^{-1}(\boldsymbol{X}^T\boldsymbol{y}+\boldsymbol{R}_d\boldsymbol{\mu}_d)$ and posterior variance $\sigma^2(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{R}_d)^{-1}$. For the example of estimating $\boldsymbol{c}^T\boldsymbol{\theta}$ with square error loss, Approach 1 leads to choosing the design to minimize the designer's expectation of the square error of the evaluator's estimate $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}_e$, which is

$$\begin{aligned} & E^{p_d(\boldsymbol{y})}E^{\pi_d(\boldsymbol{\theta}|\boldsymbol{y})}(\boldsymbol{c}^T\boldsymbol{\theta} - \boldsymbol{c}^T\hat{\boldsymbol{\theta}}_e)^2 \\ = \;& \sigma^2\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{R}_d)^{-1}\boldsymbol{c} + E^{p_d(\boldsymbol{y})}(\boldsymbol{c}^T\hat{\boldsymbol{\theta}}_d - \boldsymbol{c}^T\hat{\boldsymbol{\theta}}_e)^2 \quad . \end{aligned}$$

The first term is the designer's optimal design criterion, just as if the designer would use $\pi_d(\boldsymbol{\theta})$ for design and analysis, and the second term is the designer's expectation of the squared distance between the designer's posterior mean and the evaluator's estimate. Corresponding criteria can be derived for other loss functions. If the evaluator's prior distribution is non-informative then, just as in Approach 2, the criterion becomes that of classical non-Bayesian optimal design.

## 1.2 Linear Model Example

Consider for illustration the case of a two treatment comparison, of treatment

to control, where $\sigma^2 = 1$. The designer's prior distribution, $\pi_d(\boldsymbol{\theta})$, is such that more is known about the control mean, $\theta_1$, than the treatment mean, $\theta_2$, and that the designer believes that $\theta_2$ has a higher mean but this belief is not shared by the scientific community. Specifically, for the designer $\theta_1 \sim N(0, \frac{1}{11})$ and $\theta_2 \sim N(1,1)$ independently. The analysis, however, will be done using a less informative prior distribution, where $\theta_1 \sim N(0,1)$ and $\theta_2 \sim N(0,1)$ independently. The prior distribution for analysis is much less informative than the prior distribution for design. For estimating the difference $\theta_1 - \theta_2$ with squared error loss with $n = 10$ observations, Approach 1 gives $n_1 = 0$ and $n_2 = 10$, and, in contrast, Approach 2 gives $n_1 = 5$ and $n_2 = 5$.

Design for the designer as a single decision maker with $\pi_d(\boldsymbol{\theta})$ also leads to no observations on the control and 10 on the treatment. Neither Approach 1 nor the design for the designer as a single decision maker leads to a very satisfactory design for convincing the scientific community. No observations are taken on the control. Approach 2 leads to a design that will give the best analysis possible under the prior distribution being used for analysis.

Note that as in other examples with normal sampling, if a non-informative analysis is used for analysis, Approach 1 and Approach 2 give the same design.

## 1.3 The Non-linear Model

In a non-linear model, a normal approximation to the posterior distribution is frequently used (eg. Berger, 1985, p. 224). Suppose that each of $n$ independent variables $y_i$ is observed at an explanatory variable $x_i$, chosen from some set $\mathcal{X}$. The distribution $\boldsymbol{y}|\boldsymbol{\theta}$ depends on $\boldsymbol{\theta}$, a vector of $p$ unknown

parameters and the design $\eta$. The expected Fisher information matrix for $\boldsymbol{\theta}$ and $\eta$ is denoted by $n\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}, \eta)$.

Let $\hat{\boldsymbol{\theta}}$ denote the maximum likelihood estimate of $\boldsymbol{\theta}$. If a non-informative prior distribution is used for analysis the posterior distribution of $\boldsymbol{\theta}$ can be approximated by

$$\boldsymbol{\theta}|\boldsymbol{y}, \eta \sim N(\hat{\boldsymbol{\theta}}, [n\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\theta}}, \eta)]^{-1}). \tag{1}$$

Approach 2 in this case, leads to design optimizing some function of $\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\theta}}, \eta)$, under the designer's predictive distribution for $\hat{\boldsymbol{\theta}}$. If $\hat{\boldsymbol{\theta}}$ is a consistent estimate of $\boldsymbol{\theta}$, then the designer's predictive distribution for $\hat{\boldsymbol{\theta}}$ can be asymptotically approximated by the designer's prior distribution. Then, for example using entropy as loss our Approach 2 leads to choosing the design $\eta$ to maximize

$$\phi_1(\eta) = E^{\pi_d(\boldsymbol{\theta})} \log \det \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}, \eta) \qquad (D\text{-optimality})$$

or, using squared error as loss, leads to minimizing

$$\phi_2(\eta) = E^{\pi_d(\boldsymbol{\theta})} \operatorname{tr}(\boldsymbol{B}[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}, \eta)]^{-1}) \qquad (A\text{-optimality})$$

where the expectation is over the designer's prior distribution and $\boldsymbol{B}$ is some matrix of weights, possibly depending on $\boldsymbol{\theta}$. These are Bayesian versions of $D$-optimality and $A$-optimality for non-linear design. These criteria have become quite widely used as they are attractive alternatives to local optimality for these problems (Chernoff, 1953). Approach 2 provides further justification for them as designs using prior information for design but not for analysis. See Chaloner and Verdinelli (1995) for a review of their use.

Approach 1 would argue for different criterion involving the designer's prior precision matrix and using an approximation to the designer's posterior

distribution

$$\boldsymbol{\theta}|\boldsymbol{y},\eta \sim N(\tilde{\boldsymbol{\theta}}_d, [\boldsymbol{R}_d(\tilde{\boldsymbol{\theta}}_d) + n\boldsymbol{\mathcal{I}}(\tilde{\boldsymbol{\theta}}_d,\eta)]^{-1}) \tag{2}$$

where $\boldsymbol{R}_d(\boldsymbol{\theta})$ is the matrix of second derivatives of the logarithm of the prior density function and $\tilde{\boldsymbol{\theta}}_d$ is the generalized maximum likelihood estimate of $\boldsymbol{\theta}$ under the designer's prior distribution (Berger (1985), p. 133). This will be used is Section 5.2.

## 1.4 Summary

In summary suppose the designer has information, as a prior distribution, about the the experiment. The marginal predictive distribution of data under the designer's prior distribution, $p_d(\boldsymbol{y})$, is calculated. Approach 2 minimizes the expectation, under $p_d(\boldsymbol{y})$, of the evaluator's posterior risk. If the prior distribution for design is informative and the prior distribution for estimation is a reference or a much less informative prior distribution then this approach is that of using as much prior information as possible for design but allowing the results to "stand on their own". The results will be accepted by the scientific community. Tsutakawa (1972) also motivated this general idea and suggested using an informative prior distribution for design and a different, less informative prior distribution for inference. The designer, in this case, is also the evaluator but uses one prior distribution for design and a different, less informative, prior distribution for inference.

Approach 2 is similar to the approach of Jackson, Novick and DeKeyrel (1980) who motivate it from an adversarial situation. Using Bayesian methods to design for a frequentist analysis is also motivated and developed in Carlin and Louis (1985) and Louis and Bailey (1990) who use prior information to interpret p-values from multiple testing.

Sample size choice for both normal and Bernoulli examples are considered in Sections 2 and 3. Sample size choice for normal sampling was the primary example considered by Etzioni and Kadane (1993) using Approach 1 and is extended here to other sampling situations. Approaches 1 and 2 are compared in Section 4. The designs can be very different. A Poisson regression model is discussed in Section 5.

# 2. APPROACH 2

Denote the optimal sample size as $n_i^*$ for Approach $i, i = 1, 2$.

Our Approach 2 is first developed for any two prior distributions; that is one is not necessarily more informative than the other. Denote $R_2(n)$ to be $cn + E^{p_d(\mathbf{y})} E^{\pi_e(\theta|\mathbf{y})} L(\theta, a)$, where $a$ minimizes $E^{\pi_e(\theta|\mathbf{y})}\ L(\theta, a)$. $R_2(n)$ is the sum of the cost for sampling $n$ observations and the expectation of the evaluator's optimal Bayes risk with respect to the designer's marginal predictive distribution. Approach 2 minimizes $R_2(n)$. $R_2(n)$ is not a predictive risk in the usual sense but the sampling cost plus the expectation, under the designer's prior distribution, of the evaluator's posterior risk.

## 2.1 Normal Sampling

Consider normal sampling with mean $\theta$ and variance $\sigma^2$. Denote the mean of $n$ observations as $\bar{y}$. Let the prior distributions be normal with mean $\mu_i$ and variance $1/\tau_i, i = e, d$. Assume a constant cost $c$ per observation and without loss of generality $\sigma^2 = 1$. Under squared-error loss, the Bayes estimate of $\theta$ with respect to $\pi_e(\theta|\bar{y})$ is the posterior mean, $a = E^{\pi_e(\theta|\bar{y})}\theta$, and

$$R_2(n) = cn + \frac{1}{n + \tau_e}.$$

Minimizing $R_2(n)$ over non-negative $n$, the optimal sample size $n_2^*$ is $1/\sqrt{c} - \tau_e$ if $1/\sqrt{c} > \tau_e$; otherwise $n_2^* = 0$. Note that in this special case of normal sampling, the evaluator's posterior loss does not depend on the data, $\bar{y}$, and therefore the optimal sample size depends only on $\tau_e$, the evaluator's precision. The optimal sample size does not depend on the designer's prior distribution and is decreasing with $\tau_e$. In this special case it is also the evaluator's optimal strategy. In the case when the evaluator uses a non-informative prior distribution, $\tau_e \to 0$, the sample size converges to $1/\sqrt{c}$ which corresponds to the classical solution (see e.g. Cochran, 1977, p. 84).

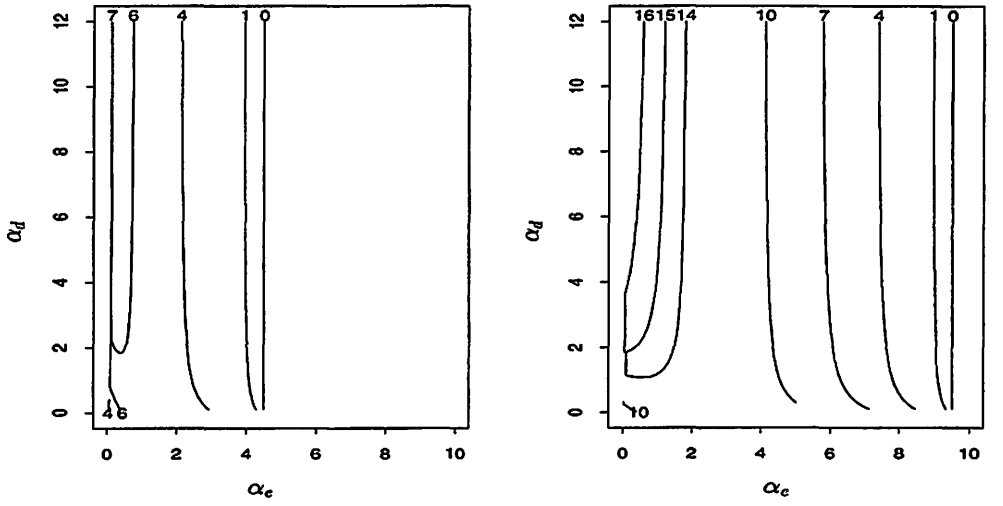
## 2.2 Bernoulli Sampling

Suppose the observations are sampled from a Bernoulli distribution with probability of success $\theta$ and Beta$(\alpha_i, \beta_i)$, $i = e, d$ are the prior distributions. Then, for all $\alpha_e, \alpha_d, \beta_e$ and $\beta_d > 0$,

$$R_2(n) = cn + \frac{n^2\alpha_d\beta_d + n(-\alpha_d\beta_d + (\alpha_d\beta_e + \alpha_e\beta_d)(\alpha_d + \beta_d + 1)) + \alpha_e\beta_e(\alpha_d + \beta_d)(\alpha_d + \beta_d + 1)}{(\alpha_d + \beta_d)(\alpha_d + \beta_d + 1)(n + \alpha_e + \beta_e)^2(n + \alpha_e + \beta_e + 1)}$$

Unlike the case with a normal likelihood, $R_2(n)$ is hard to simplify and there is no closed form expression for the optimal sample size $n_2^*$. In the special case $\alpha_e = k\beta_e$, $k > 0$, as $\alpha_e$ goes to $+\infty$ then $R_2(n)$ tends to $cn$ which does not depend on $\alpha_d$ and $\beta_d$. The evaluator's prior opinion is very informative and so the evaluator's posterior loss will be close to zero, whatever the data. The optimal sample size is therefore 0.

For $\alpha_e = \beta_e$ and $\alpha_d = \beta_d$, Figure 2.1(a) and Figure 2.1(b) plot the contours of $n_2^*$ as a function of $\alpha_e$ and $\alpha_d$ for cost $= 0.0025$ and $0.000625$. The region to the right of line 0 is the region of $n_2^* = 0$ and the sample sizes




(a) (b)

Figure 2.1: The contours of $n_2^*$ as a function of $\alpha_e$ and $\alpha_d$ for $\alpha_e = \beta_e$, $\alpha_d = \beta_d$ and (a) cost = 0.0025, (b) cost = 0.000625.

between line 0 and line 1 are all 1. Table 2.1 reports some numerical results for cost = 0.0025, and both prior means are $\frac{1}{2}$. The sample size does not depend much on $\alpha_d$ and $\beta_d$.

The case $\alpha_e \to 0$, $\beta_e \to 0$ is of special interest as it corresponds to vague prior information for estimation. The sample size $n$ should be restricted to at least two. If $n = 0$ or $n = 1$ then $\pi_e(\theta|\bar{y})$ is improper and thus $R_2(n)$ is undefined. The optimal sample sizes for $\alpha_e \to 0$, $\beta_e \to 0$ and cost = 0.0025 are shown in Table 2.2. In this case, the sample size is maximized at $\alpha_d = \beta_d$ in each row and each column. If $\alpha_d = \beta_d = \alpha$, the sample size increases with $\alpha$ and is bounded by 7. (With cost = 0.0025, $n_2^*$ is 7 for $\alpha_d = \beta_d = 10000$.)

# 3. APPROACH 1

Etzioni and Kadane (1993) consider the same normal sampling problem for a normal likelihood with unknown mean $\theta$ and known variance $\sigma^2$. The goal of their approach is to find the optimal sample size $n_1^*$ for which the designer's predictive risk of sampling $n$ observations for the evaluator's posterior estimate is minimized. Their results for normal sampling are summarized here and their approach is extended here to Bernoulli sampling. Both approaches can also be extended to Poisson sampling (see Tsai, 1999).

## 3.1 Normal Sampling

Assume $\pi_i(\theta)$ to be normal, $N(\mu_i, 1/\tau_i), i = e, d$. Let $\sigma^2$ be 1. Assume a constant cost $c$ per observation and squared-error loss. Let $r = \tau_e(\tau_e - \tau_d)/\tau_d + \tau_e^2(\mu_d - \mu_e)^2$, $y = \sqrt{c}(n + \tau_e)$, $\tilde{r} = \sqrt{c}r$, $g(y) = y + 1/y + \tilde{r}/y^2$, $\tilde{n} = \sqrt{c}n$ and $\tilde{\tau}_e = \sqrt{c}\tau_e$. Then the designer's predictive risk of sampling $n$ observations for the evaluator's estimate, $R_1(n)$, is

$$\begin{aligned} R_1(n) &= cn + \frac{1}{n + \tau_e} + \frac{r}{(n + \tau_e)^2} \\ &= \sqrt{c}g(y) - c\tau_e. \end{aligned}$$

**Theorem 1 [Etzioni and Kadane, 1993].** *Let $y(\tilde{r})$ be the largest real root of the cubic $y^3 - y - 2\tilde{r} = 0$. Then the optimal sample size is $n_1^* = \tilde{n}_1^*/\sqrt{c}$ where*

$$\tilde{n}_1^* = \begin{cases} y(\tilde{r}) - \tilde{\tau}_e & \text{if } \tilde{r} > (-\frac{1}{3\sqrt{3}}) \text{ and } -\frac{\tilde{r}}{y^2(\tilde{r})} < \tilde{\tau}_e < y(\tilde{r}), \\ 0 & \text{otherwise.} \end{cases}$$

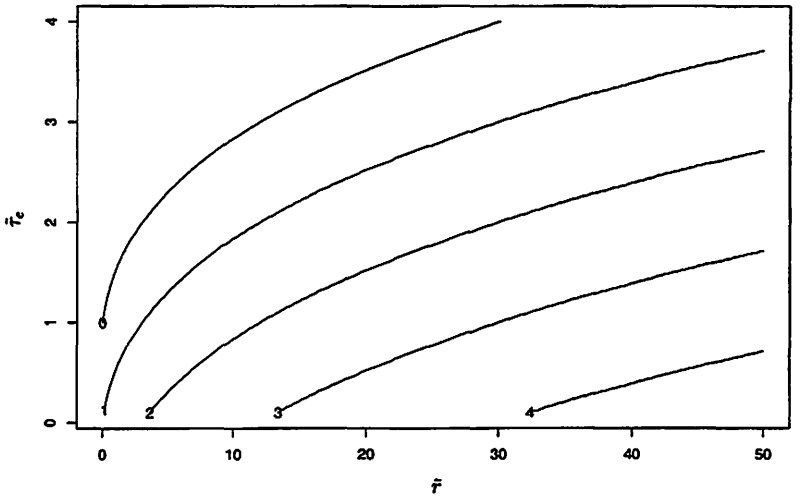



Figure 3.1: The contours of $\tilde{n}_1^*$ as a function of $\tilde{r}$ and $\tilde{\tau}_e$.

Figure 3.1 plots contours of $\tilde{n}_1^*$ as a function of $\tilde{r}$ and $\tilde{\tau}_e$.

From Theorem 1, it is optimal not to sample when the prior precisions for both estimation and design are high and the two prior means are close to each other. When the prior means are far apart or the prior precision for design is small, more observations will be needed.

For the case $|\mu_d - \mu_e|$ and $\tau_d$ fixed, $\tilde{r}$ tends to 0 as $\tilde{\tau}_e$ goes to 0. This implies $\tilde{n}_1^*$ converges to 1. Thus the optimal sample size, $n_1^*$, converges to $1/\sqrt{c}$ which is the sample size under a flat prior and it also corresponds to the classical result.

When the two prior distributions agree, the optimal sample size will be $1/\sqrt{c} - \tau$ if $\tau < 1/\sqrt{c}$, otherwise $n_1^* = 0$.

## 3.2 Bernoulli Sampling

Choosing the sample size for Bernoulli observations with probability of suc-

cess $\theta$ was not considered by Etzioni and Kadane. Let the prior distributions used for estimation and design be the conjugate Beta distributions, Beta$(\alpha_e, \beta_e)$ and Beta$(\alpha_d, \beta_d)$ with strictly positive parameters $\alpha_e$, $\beta_e$, $\alpha_d$ and $\beta_d$.

Let $cn$ be the sampling cost, and

$$\begin{aligned} y &= \sqrt{c}(n+\alpha_e+\beta_e)\left(\frac{(\alpha_d+\beta_d)(\alpha_d+\beta_d+1)}{\alpha_d\beta_d}\right)^{1/2} \\ r &= \sqrt{c}\left(\frac{(\alpha_d+\beta_d)(\alpha_d+\beta_d+1)}{\alpha_d\beta_d}\right)^{3/2} \times \\ &\quad \frac{(\alpha_e\beta_d-\alpha_d\beta_e)^2(\alpha_d+\beta_d+1)+\alpha_d\beta_d(\alpha_e+\beta_e)((\alpha_e+\beta_e)-(\alpha_d+\beta_d))}{(\alpha_d+\beta_d)^2(\alpha_d+\beta_d+1)} \end{aligned}$$

and

$$g(y) = y+\frac{1}{y}+\frac{r}{y^2}$$

$$\lambda = \sqrt{c}(\alpha_e+\beta_e)\sqrt{\frac{(\alpha_d+\beta_d)(\alpha_d+\beta_d+1)}{\alpha_d\beta_d}}.$$

Under squared-error loss,

$$R_1(n) = \sqrt{c}\sqrt{\frac{\alpha_d\beta_d}{(\alpha_d+\beta_d)(\alpha_d+\beta_d+1)}}g(y) - c(\alpha_e+\beta_e).$$

Let $y(r)$ denote the largest real root of the cubic $y^3 - y - 2r = 0$. By Theorem 1, the optimal sample size for Bernoulli sampling is

$$n_1^* = \begin{cases} \frac{\alpha_e+\beta_e}{\lambda}y(r) - (\alpha_e+\beta_e) & \text{if } r > -\frac{1}{3\sqrt{3}} \text{ and } -\frac{r}{y^2(r)} < \lambda < y(r), \\ 0 & \text{otherwise.} \end{cases}$$

Consider the case $\alpha_e/\beta_e = \alpha_d/\beta_d = k$ with $k > 0$ first. The designer and the evaluator have the same prior mean but different prior variances. In

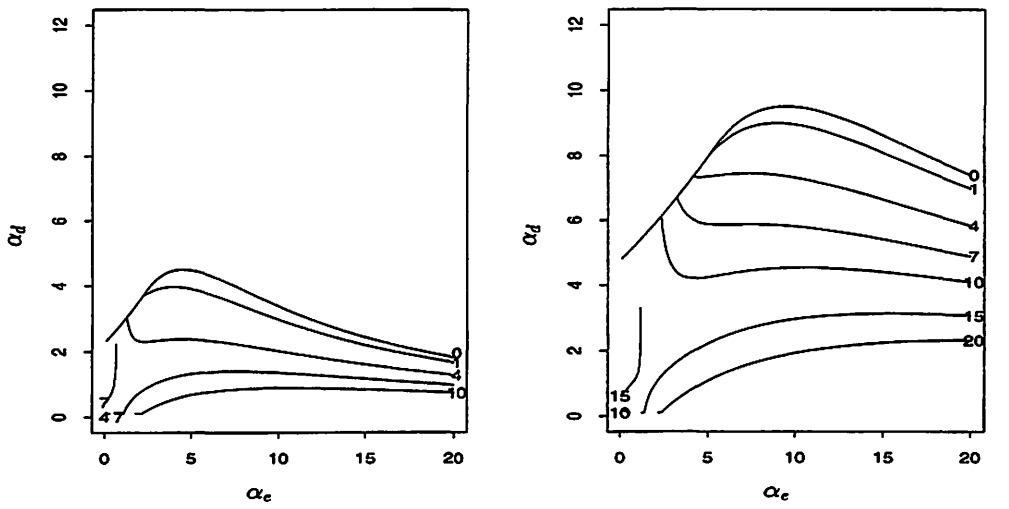



Figure 3.2: The contours of $n_1^*$ as a function of $\alpha_e$ and $\alpha_d$ for $\alpha_e = \beta_e$, $\alpha_d = \beta_d$ and (a) cost = 0.0025, (b) cost = 0.000625.

this case, the optimal sample size is 0 for large $\alpha_d$. If $\alpha_e$ is fixed, then $n_1^*$ decreases as $\alpha_d$ increases for $\alpha_d < \alpha_e$.

For $k = 1$, both the prior distributions are symmetric about 1/2 and the prior precisions are $4(2\alpha_e + 1)$ and $4(2\alpha_d + 1)$. Figure 3.2(a) and Figure 3.2(b) plot contours of $n_1^*$ as a function of $\alpha_e$ and $\alpha_d$ for cost = 0.0025 and 0.000625. The sample size is 0 for any $(\alpha_e, \alpha_d)$ lies in the region above line 0 and the region between line 0 and line 1 is the region of $n_1^* = 1$. The numerical results for cost = 0.0025 and 0.000625 are summarized in Table 3.1 and Table 3.2. The optimal sample sizes are small when the designer has high prior precision. Moreover, if there is a single prior distribution with mean $\frac{1}{2}$, let $\alpha = \alpha_e = \alpha_d$, for $\alpha \geq (-1 + \sqrt{1 + 1/c})/4$ it is optimal not to sample. Table 3.3 summarizes the sample sizes with different $\alpha$'s and costs.

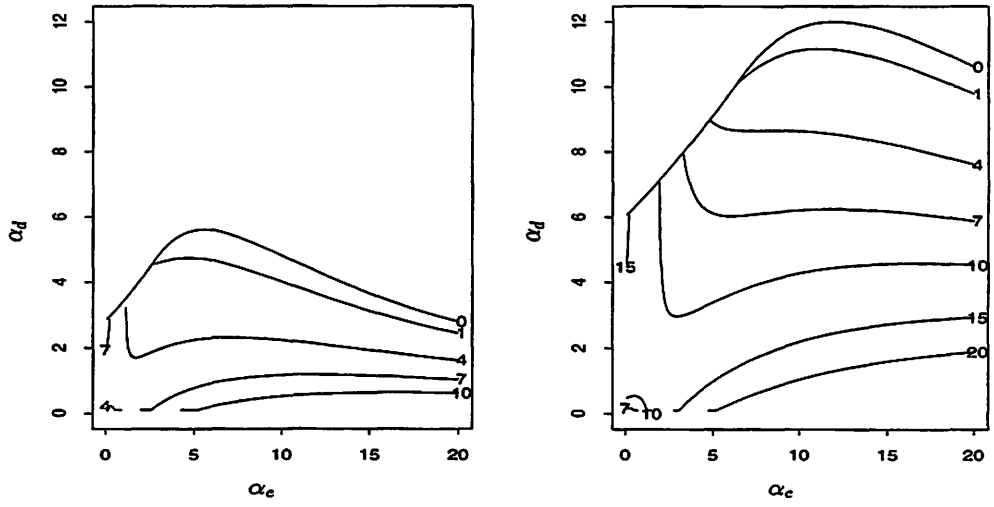



Figure 3.3: The contours of $n_1^*$ as a function of $\alpha_e$ and $\alpha_d$ for $\alpha_e = 4\beta_e$, $\alpha_d = 4\beta_d$ and (a) cost = 0.0025, (b) cost = 0.000625.

It's not surprising that the sample size decreases with cost. For a small value of cost, say 0.0001, the sample size changes from 7 to 20 when $\alpha$ increases from 0.01 to 0.1.

For $k = 4$, Figure 3.3(a) and Figure 3.3(b) plot contours of $n_1^*$ as a function of $\alpha_e$ and $\alpha_d$ with $\alpha_i = 4\beta_i$, $i = e, d$ for cost = 0.0025 and 0.000625. The results are similar to the results for $k = 1$. In the case $\alpha_e = \alpha_d$, the optimal sample size is 0 if $\alpha_e \geq 2\left(-1 + \sqrt{1 + 16/(25c)}\right)/5$.

If both $\alpha_e$ and $\beta_e$ go to 0, similar to the conditions in Approach 2, the sample size should be restricted to at least 2 so that the evaluator's posterior mean and variance are well-defined. If $\alpha_e$ and $\beta_e$ tend to 0 and $\alpha_d$ and $\beta_d$ are fixed, then $n_1^* = \max\{2, \sqrt{\alpha_d\beta_d}/\sqrt{c(\alpha_d + \beta_d)(\alpha_d + \beta_d + 1)}\}$. In this case, more data are required if both the parameters $\alpha_d$ and $\beta_d$ are small or if both

are large. Moreover, $n_1^*$ approaches 2 from above as one of the parameters goes to $+\infty$ with the other one fixed. If $c < \frac{1}{2}$ and $\alpha_d$ and $\beta_d$ are equal and go to $+\infty$ then $n_1^*$ approaches $1/(2\sqrt{c})$ from below.

When the two prior distributions agree and are proper, $n_1^*$ is the maximum of $\sqrt{\alpha\beta}/\sqrt{c(\alpha+\beta)(\alpha+\beta+1)} - (\alpha+\beta)$ and 0.

# 4. COMPARISON OF TWO APPROACHES

Recall that the optimal sample size is denoted as $n_i^*$ for Approach $i, i = 1, 2$.

## 4.1 Normal Example

Suppose that observations $y_i|\theta$ are independent observations from a $N(\theta, 1)$ distribution. Observations cost $c = 0.0025$ units and the loss is square error. The prior distributions are $\pi_d(\theta)$ is $N(0, 1/20)$ and $\pi_e(\theta)$ is $N(0, 1)$: the prior distribution for design is much more precise than the prior distribution for estimation. Approach 2 gives $n_2^* = 19$, this minimizes the evaluator's posterior loss. Approach 1, minimizing the designer's expectation of the loss for the suboptimal estimator that the evaluator will use, gives $n_1^* = 0$, take no observations. For Approach 1 this makes sense as the evaluator's estimate, if no observations will be taken, is zero, which is a good estimator under the designer's very precise prior distribution. Approach 1 takes no account of the evaluator's posterior variance, which will remain large.

## 4.2 Normal Sampling Overview

Under a normal likelihood with known variance 1, the optimal sample size

$n_1^*$ is $y(\tilde{r})/\sqrt{c} - \tau_e$ if both $\tilde{r} > -1/(3\sqrt{3})$ and $-\tilde{r}/y^2(\tilde{r}) < \tilde{\tau}_e < y(\tilde{r})$; and is 0 otherwise. Approach 2 gives $n_2^* = \max\{0, 1/\sqrt{c} - \tau_e\}$. Approach 1 and Approach 2 give the same sample size if $r$ equals 0. Thus in the case of a non-informative prior distribution for estimation, that is $\tau_e \to 0$, both $n_1^*$ and $n_2^*$ converge to $1/\sqrt{c}$ when $|\mu_e - \mu_d|$ and $\tau_d$ are fixed. In addition, when the two prior distributions agree, so do $n_1^*$ and $n_2^*$, and they are both equal to $\max\{0, 1/\sqrt{c} - \tau_e\}$. In general, if $r$ is greater than 0, more observations are needed under Approach 1. That is, if $\tau_e$ is fixed, then $n_1^* \geq n_2^*$ when the two prior means are far apart or the designer's precision is less than the evaluator's precision. Alternatively, if $r$ is less than 0, then $n_1^* \leq n_2^*$.

## 4.3 Bernoulli Example

Consider a similar example to that of 4.1 but with Bernoulli sampling. That is suppose that observations $y_i|\theta$ are independent observations from a Bernoulli distribution with mean $\theta$.. Observations cost $c = 0.0025$ units and the loss is square error. The prior distributions are $\pi_d(\theta)$ is Beta(10, 10) and $\pi_e(\theta)$ is uniform, Beta(1, 1). Again, the prior distribution for design is much more precise than the prior distribution for estimation. Approach 2 gives $n_2^* = 6$, this minimizes the evaluator's posterior loss. Approach 1, gives $n_1^* = 0$, take no observations. Again for Approach 1 this makes sense as if no observations will be taken $\hat{\theta}_e = 0$, which is a good estimator under the designer's very precise prior distribution.

## 4.4 Bernoulli Sampling Overview

For a normal sample with variance 1, after re-parameterization, the optimal sample size is a function of $\tilde{\tau}_e$ and $\tilde{r}$ in Approach 1. The relationship between the optimal sample size and those two parameters, $\tilde{\tau}_e$ and $\tilde{r}$ can be

easily seen in a contour plot. Under Bernoulli sampling, however, the optimal sample size can not be re-parameterized as a function of two parameters. Only some special cases are discussed here. In the case $\alpha_e = \beta_e$ and $\alpha_d = \beta_d$, the evaluator and the designer have the same prior mean of $\frac{1}{2}$ but different prior variances. If the designer's precision is high the optimal sample size for Approach 1 is 0. If the evaluator has high precision Approach 2 has an optimal sample size of 0. For $\alpha_d$ and $\beta_d$ fixed, if both $\alpha_e$ and $\beta_e$ go to 0, $R_1(n)$ tends to $cn + \alpha_d\beta_d/(n(\alpha_d+\beta_d)(\alpha_d+\beta_d+1))$ and $R_2(n)$ converges to $cn + (n-1)\alpha_d\beta_d/(n(n+1)(\alpha_d+\beta_d)(\alpha_d+\beta_d+1))$, which is similar, but not quite identical. When the two prior distributions agree and are proper, $n_1^*$ is $\max\left\{0, \sqrt{\alpha\beta}/\sqrt{c(\alpha+\beta)(\alpha+\beta+1)} - (\alpha+\beta)\right\}$, the same as $n_2^*$.

## 5. POISSON OPTIMAL DESIGN EXAMPLE

For a special case of Poisson sampling, Approach 2 can be implemented exactly, under squared-error loss, without the use of approximations. Consider a model where observations $y_{ij}$ are from Poisson distributions with mean $\theta x_i$ where $\theta$ is a unknown parameter and $x_i$ are the design points which are non-negative and bounded above by a positive value $B$. Given an overall sample size $n$, suppose $n_i$ observations are from Poisson $(\theta x_i)$ for $i = 1, 2, \cdots, k$, $k \leq n$ and $\sum_{i=1}^{k} n_i = n$. Note that $T = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$ is a sufficient statistic for $\theta$. For Approach 2, the purpose of the experiment is to choose $x_i$, $n_i$ and $k$ such that $E^{p_d(\sum Y_{ij}, \boldsymbol{x})} E^{\pi_e(\theta|\sum Y_{ij}, \boldsymbol{x})} L(\theta, a)$ is minimized where $a$ minimizes $E^{\pi_e(\theta|\sum Y_{ij}, \boldsymbol{x})} L(\theta, a)$.

Let the prior distribution of $\theta$ be a gamma distribution, $\pi_i(\theta) = \beta_i^{\alpha_i}\theta^{\alpha_i - 1}$ $\exp(-\beta_i\theta)/\Gamma(\alpha_i)$, $i = e, d$. Then given $\theta$ and $\boldsymbol{x}$, $T$ is Poisson with mean

$\theta \sum_{i=1}^{k} n_i x_i$. The posterior distributions of $\theta$ given $T$ and $\boldsymbol{x}$ are

$$\pi_i(\theta|T,\boldsymbol{x}) = \frac{(\beta_i + \sum_{i=1}^{k} n_i x_i)^{\alpha_i+T}}{\Gamma(\alpha_i+T)}\theta^{\alpha_i+T-1}\exp\left(-\left(\beta_i + \sum_{i=1}^{k} n_i x_i\right)\theta\right), i = e, d.$$

In this section first $n$ will be considered fixed and the designer's expectation of the evaluator's posterior expected loss will be minimized then second, for fixed cost $c$ per observation the best choice of $n$ will be found.

## 5.1 Quadratic Loss for Approach 2

Under squared-error loss, for fixed $n$,

$$E^{p_d(T|\boldsymbol{x})}\left[Var^{\pi_e(\theta|T,\boldsymbol{x})}\theta\right] = \frac{\alpha_e\beta_d + \alpha_d\sum_{i=1}^{k} n_i x_i}{\beta_d(\beta_e + \sum_{i=1}^{k} n_i x_i)^2}$$

which depends on $\sum_{i=1}^{k} n_i x_i$.

Let $M = \sum_{i=1}^{k} n_i x_i$ and, for fixed $n$, denote $\phi(M)$ to be $(\alpha_e\beta_d+\alpha_d M)/(\beta_d(\beta_e+M)^2)$. Taking first and second derivatives of $\phi(M)$ with respect to $M$:

$$\begin{aligned}\frac{\partial}{\partial M}\phi(M) &= \frac{1}{\beta_d}(\beta_e + M)^{-3}(\alpha_d\beta_e - 2\alpha_e\beta_d - \alpha_d M)\\ \frac{\partial^2}{\partial M^2}\phi(M) &= \frac{1}{\beta_d}(\beta_e + M)^{-4}(-4\alpha_d\beta_e + 6\alpha_e\beta_d + 2\alpha_d M).\end{aligned}$$

Note that $M$ is non-negative and bounded above by $nB$ and below by 0. If $\alpha_d\beta_e - 2\alpha_e\beta_d \geq 0$, then $\frac{\partial}{\partial M}\phi(M) = 0$ at $M_0 = (\alpha_d\beta_e - 2\alpha_e\beta_d)/\alpha_d$ and $\frac{\partial^2}{\partial M^2}\phi(M)$ is negative at $M = M_0$ . The function $\phi(M)$ is concave with maximum $\phi(M_0)$ and minimum $\min\{\phi(0), \phi(nB)\}$. If $\alpha_d\beta_e - 2\alpha_e\beta_d < 0$, then $\frac{\partial}{\partial M}\phi(M) < 0$ for all $M \geq 0$, and $\phi(M)$ is decreasing with $M$. So $\phi(M)$ is minimized at $M = nB$.

For fixed $n$, therefore, if $\alpha_d\beta_e - 2\alpha_e\beta_d \geq 0$ and $\phi(0) < \phi(nB)$ then it is optimal to put all observations at $x = 0$; otherwise the optimal design puts

all observations at $x = B$. In both cases the optimal design is a one-point design.

It is also clear that if all observations are at $x = 0$ then $n^* = 0$.

If the optimal design puts all observations at $x = B$, then let $r = \sqrt{c/B}\sqrt{\beta_d/\alpha_d}^3(\alpha_e\beta_d - \alpha_d\beta_e)/\alpha_d$ and $y(r)$ denote the largest real root of the cubic $y^3 - y - 2r = 0$. Then by Theorem 1,

$$n^* = \begin{cases} \sqrt{\frac{B}{c}\frac{\alpha_d}{\beta_d}}y(r) - \beta_e & \text{if } r > -\frac{1}{3\sqrt{3}} \text{ and } -\frac{r}{y^2(r)} < \sqrt{\frac{c\beta_d}{B\alpha_d}}\beta_e < y(r), \\ 0 & \text{otherwise.} \end{cases}$$

## 5.2 Logarithmic Loss for Approach 2

When the loss function is $-\log(a)$, $E^{\pi_e(\theta|\sum_{i=1}^k\sum_{j=1}^{n_i}Y_{ij},\boldsymbol{x})}(-\log(a))$ is minimized at $a = \pi_e(\theta|\sum_{i=1}^k\sum_{j=1}^{n_i}Y_{ij},\boldsymbol{x})$ (Bernardo, 1979). The posterior distribution of $\theta$ and the distribution of the sufficient statistic $\sum_{i=1}^k\sum_{j=1}^{n_i}Y_{ij}$ again depend on the design only through $M = \sum_{i=1}^k n_i x_i$. Let $h$ be $\beta_d/(\beta_d + M)$. Then the designer's predictive distribution of the data is

$$\begin{aligned} p_d(t|M) &= \int p(t|\theta, M)\pi_d(\theta)d\theta \\ &= \frac{\Gamma(\alpha_d + t)}{\Gamma(\alpha_d)t!}h^{\alpha_d}(1-h)^t. \end{aligned}$$

Taking the first and second derivatives of $\log p(t|\theta, M)$ and $\log \pi_e(\theta)$,

$$\frac{\partial}{\partial\theta}\log p(t|\theta, M) = \frac{t}{\theta} - M; \qquad \frac{\partial^2}{\partial\theta^2}\log p(t|\theta, M) = -\frac{t}{\theta^2}$$

and

$$\frac{\partial}{\partial\theta}\log \pi_e(\theta) = \frac{\alpha_e - 1}{\theta} - \beta_e; \qquad \frac{\partial^2}{\partial\theta^2}\log \pi_e(\theta) = -\frac{\alpha_e - 1}{\theta^2}$$

The generalized maximum likelihood estimate of $\theta$, $\hat{\theta}$, is $(t+\alpha_e-1)/(\beta_e+M)$. Therefore by using approximation (2), if $t+\alpha_e-1>0$ then the evaluator's posterior distribution can be approximated by

$$\pi_e(\theta|t,M) \approx N\left(\frac{t+\alpha_e-1}{\beta_e+M}, \frac{t+\alpha_e-1}{(\beta_e+M)^2}\right).$$

Minimizing $E^{p_d(t|M)}E^{\pi_e(\theta|t,M)}(-\log\pi_e(\theta|t,M))$ over $M$ is equivalent to minimizing $\phi(M) = -2\log(\beta_e+M)+E^{p_d(t|M)}(\log(t+\alpha_e-1))$. Table 5.1 gives the results of some numerical minimizations of $\phi(M)$ for $nB = 100$. In all cases the optimal design was a one-point design putting all the observations at $x = 0$ or all at $x = B$. When it is optimal to take all the observations at $x = 0$ it can be shown that the optimal sample size, $n^*$ must be 0. When all observations are taken at $x = B$ it is not possible to find a closed form for $n^*$.

## 5.3 Approach 1

The posterior distribution $\pi_i(\theta|t,M)$ can be approximated by a normal distribution with mean $(t+\alpha_i-1)/(\beta_i+M)$ and variance $(t+\alpha_i-1)/(\beta_i+M)^2$ where $t+\alpha_i-1>0, i=e,d$. Under the logarithmic loss function, for fixed $M$,

$$\begin{aligned} & E^{p_d(t|M)}E^{\pi_d(\theta|t,M)}\left[-\log(\pi_e(\theta|t,M))\right] \\ \approx\ & E^{p_d(t|M)}\left[-\log(\beta_e+M)+\frac{1}{2}\log(2\pi(t+\alpha_e-1))\right. \\ & \left.+\frac{(\beta_e+M)^2}{2(t+\alpha_e-1)}\left(\frac{t+\alpha_d-1}{(\beta_d+M)^2}+\left(\frac{t+\alpha_d-1}{\beta_d+M}-\frac{t+\alpha_e-1}{\beta_e+M}\right)\right)\right]. \end{aligned}$$

Numerical results for some $\alpha_e$, $\beta_e$, $\alpha_d$, $\beta_d$ and $nB = 100$ are summarized in Table 5.2. The optimal design is again the one-point design.

## 5.4 Discussion

Approach 1 and Approach 2 give different designs in the examples. For example, if $\beta_e$ approaches 0, $\alpha_d = \beta_d = 10$ and $\alpha_e - 1 = 0.5$, then the optimal design is at $x = 0$ for Approach 1, but is at $x = B$ for Approach 2.

In the special case where the two prior distributions agree, $\pi_e(\theta) = \pi_d(\theta)$, the two approaches give the same design. Let $\alpha_e = \alpha_d = \alpha$ and $\beta_e = \beta_d = \beta$. Under squared-error loss, $\phi(M) = \alpha/(\beta(\beta+M))$ which is decreasing with $M$. Therefore, $\phi(M)$ is minimized at $M = nB$. When a logarithmic loss function is used, $\phi(M) = -2\log(\beta + M) + E^{p(t|M)}(\log(t + \alpha_e - 1))/2 + (\log(2\pi))$ which must be calculated numerically. The optimal design is at $x = B$ for $\alpha = 1.1, 2, 10$, $\beta = 0.5, 1, 10$ and $nB = 100$. When $\alpha = 1.1$ and $\beta = 100$, $\phi(M)$ is minimized at $M = 0$.

# 6. CONCLUSION

Approach 1 gives the design that minimizes the designer's own risk for the evaluator's suboptimal estimate. But in many situations, experiments are designed with a goal of submitting the results for publication. The published analysis will typically correspond to a non-informative or much less informative prior distribution, or the analysis may be by maximum likelihood or some other non-Bayesian methods. Approach 2 reflects this situation and the goal is to design such that the inferences reported are as precise as possible. This approach reflects the fact that an experimenter may have substantial information that he or she is willing to use for design but would prefer to use a non-informative prior for inference so that the results will be widely believed. This situation is likely to occur frequently and complements Approach 1.

The strict, normative, Bayesian theory implies that if prior information is available it should be used for both design and analysis. This can lead to designs, for example, where few, or no, observations are taken on control groups, because there is a lot of prior information about the control treatment. The strict normative analysis will subsequently depend on that prior information but be less convincing to a wider audience who will not share that prior information. Approach 2 formalizes a way of using prior information for design but not for analysis.

# Tables

Table 2.1: The Optimal Sample Sizes for Bernoulli Distribution with $\alpha_e = \beta_e$, $\alpha_d = \beta_d$, cost $= 0.0025$ for Approach 2

| | | $\alpha_d$ | | | | |
|---|---|---|---|---|---|---|
| | | 0.0001 | 0.5 | 1 | 2 | 10 |
| | 0.5 | 6 | 6 | 6 | 7 | 7 |
| | 1 | 6 | 6 | 6 | 6 | 6 |
| | 2 | 6 | 5 | 5 | 5 | 5 |
| $\alpha_e$ | 3 | 5 | 4 | 4 | 3 | 3 |
| | 4 | 3 | 2 | 2 | 2 | 1 |
| | 5 | 0 | 0 | 0 | 0 | 0 |
| | 10 | 0 | 0 | 0 | 0 | 0 |

Table 2.2: The Optimal Sample Sizes for Bernoulli Distribution with $\alpha_e \to 0$, $\beta_e \to 0$ and cost=0.0025 for Approach 2

| | | $\beta_d$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0.5 | 1 | 2 | 3 | 4 | 7 | 10 |
| | 0.5 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| $\alpha_d$ | 3 | 2 | 2 | 2 | 7 | 7 | 2 | 2 |
| | 4 | 2 | 2 | 2 | 7 | 7 | 7 | 2 |
| | 7 | 2 | 2 | 2 | 2 | 7 | 7 | 7 |
| | 10 | 2 | 2 | 2 | 2 | 2 | 7 | 7 |

Table 3.1: The Optimal Sample Sizes for Bernoulli Distribution with $\alpha_e = \beta_e$, $\alpha_d = \beta_d$ and cost $= 0.0025$ for Approach 1

| | | $\alpha_d$ | | | | |
|---|---|---|---|---|---|---|
| | | 0.5 | 1 | 3 | 5 | 10 |
| | 0.5 | 6 | 7 | 0 | 0 | 0 |
| | 1 | 7 | 6 | 0 | 0 | 0 |
| | 2 | 8 | 7 | 3 | 0 | 0 |
| $\alpha_e$ | 3 | 10 | 8 | 3 | 0 | 0 |
| | 4 | 11 | 8 | 3 | 0 | 0 |
| | 5 | 12 | 9 | 3 | 0 | 0 |
| | 10 | 14 | 10 | 1 | 0 | 0 |

Table 3.2: The Optimal Sample Sizes for Bernoulli Distribution with $\alpha_e = \beta_e$, $\alpha_d = \beta_d$ and cost $= 0.000625$ for Approach 1

| | | $\alpha_d$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.5 | 1 | 3 | 5 | 8 | 10 |
| | 0.5 | 13 | 15 | 17 | 17 | 0 | 0 |
| | 1 | 14 | 14 | 15 | 15 | 0 | 0 |
| | 2 | 17 | 15 | 13 | 12 | 0 | 0 |
| $\alpha_e$ | 3 | 20 | 17 | 13 | 10 | 0 | 0 |
| | 4 | 23 | 19 | 13 | 9 | 0 | 0 |
| | 5 | 25 | 21 | 13 | 9 | 0 | 0 |
| | 10 | 35 | 28 | 15 | 9 | 3 | 0 |

Table 3.3: The Optimal Sample Sizes for Bernoulli Distribution with $\alpha_e = \alpha_d = \beta_e = \beta_d = \alpha$ for Approach 1

| | | cost/10000 | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 10 | 20 | 30 | 40 | 50 |
| | 0.01 | 7 | 2 | 2 | 1 | 1 | 1 |
| | 0.05 | 15 | 5 | 3 | 3 | 2 | 2 |
| | 0.1 | 20 | 6 | 4 | 4 | 3 | 3 |
| $\alpha$ | 0.5 | 34 | 10 | 7 | 5 | 5 | 4 |
| | 1 | 39 | 11 | 7 | 5 | 4 | 4 |
| | 5 | 38 | 5 | 1 | 0 | 0 | 0 |
| | 10 | 29 | 0 | 0 | 0 | 0 | 0 |

Table 5.1: The Poisson Optimal Design for Approach 2

| $\alpha_d$ | $\beta_d$ | $\alpha_e - 1 = 0.5, 100$ $\beta_e = 0, 0.9, 1$ | $\alpha_e - 1 = 0.5$ $\beta_e = 100$ | $\alpha_e - 1 = 100$ $\beta_e = 100$ |
|---|---|---|---|---|
| 0.1 | 1,10 | | | |
| 0.1,1 | 50 | $B$ | $B$ | $B$ |
| 0.1,1 | 100 | | | |
| 0.1,10 | 0.1 | | | |
| 1 | 1 | $B$ | 0 | $B$ |
| 1,10 | 10 | | | |
| 10,50,100 | 50,100 | | | |
| 1,50 | 0.1 | | | |
| 10 | 1 | $B$ | 0 | 0 |
| 50,100 | 1,10 | | | |

Table 5.2: The Poisson Optimal Design for Approach 1

| | | $\alpha_e - 1 = 0.5$ | | | | $\alpha_e - 1 = 100$ |
|---|---|---|---|---|---|---|
| $\alpha_d$ | $\beta_d$ | $\beta_e = 0$ | $\beta_e = 0.9$ | $\beta_e = 1$ | $\beta_e = 100$ | $\beta_e = 0, 0.9, 1, 100$ |
| 0.1<br>1 | 1<br>1,10,50,100 | $B$ | $B$ | $B$ | 0 | $B$ |
| 0.1 | 10,50,100 | $B$ | $B$ | $B$ | $B$ | $B$ |
| 10<br>50 | 10<br>50 | 0 | $B$ | $B$ | $B$ | $B$ |
| 50,100 | 100 | 0 | 0 | $B$ | $B$ | $B$ |

# References


Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.

Bernardo, J. M. (1979), "Expected information as expected utility," *The Annals of Statistics*, **7**, 686-690.

Burt, J. (1990), "Towards Agreement: Bayesian Experimental Design," Technical Report 90-41, Purdue University, Department of Statistics.

Carlin, J. B. and Louis, T. A. (1985), "Controlling error rates by using conditional expected power to select tumor sites". *Proceedings of the 1985 Biopharmaceutical Section Proceedings of the American Statistical Association*, 11-16.

Chaloner, K. and Verdinelli, I. (1995), "Bayesian experimental design: a review," *Statistical Science*, **10**, 273-304.

Chernoff, H. (1953), "Locally optimal designs for estimating parameters," *Ann. Math. Statist.* **24**, 586-602.

Cochran, W. G. (1977), *Sampling Techniques*, Third Edition, John Wiley and Sons, New York.

Etzioni, R. and Kadane, J. B. (1993), "Optimal experimental design for another's analysis," *Journal of the American Statistical Association*, **88**, 1404-1411.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

Jackson, P. H., Novick, M. R. and DeKeyrel, D. F. (1980), "Adversary preposterior analysis for simple parametric Models," In Arnold Zellner and Harold Jeffreys, editors, *Bayesian Analysis in Econometrics and Statistics,* Chapter 10, 113-132, North-Holland, Amsterdam.

Lindley, D. V. and Singpurwalla, N. D. (1991). "On the evidence needed to reach agreed action between adversaries, with application to acceptance sampling," *Journal of the American Statistical Association* 86, 933–937.

Lindley, D. V. and Singpurwalla, N. D. (1993). "Adversarial Life Testing," *Journal of the Royal Statistical Society*, B, 55, 837–847.

Lodh, M. (1993). "Experimental Studies by Distinct Designer and Estimators," PhD Thesis, Department of Statistics, Carnegie Mellon University.

Louis, T. A. and Bailey, J. K. (1990). "Controlling error rates using prior information and marginal totals to select tumor sites", *Journal of Statistical Planning and Inference*, **24**, 297–316.

Pukelsheim, F. (1993), *Optimal Design of Experiments,* Wiley, New York.

Silvey, S. D. (1980), *Optimal Design,* Chapman and Hall, London.

Tsai, C. P. (1999). University of Minnesota PhD Thesis, *in preparation.*

Tsutakawa, R. K. (1972), "Design of Experiment for Bioassay," *Journal of the American Statistical Association*, **67**, 584-590.